2008

Using GIS to depict resource risk from probable cannabis cultivation sites

Cammie d'Entremont Partelow
San Jose State University

USING GIS TO DEPICT RESOURCE RISK FROM PROBABLE *CANNABIS* CULTIVATION SITES

A Thesis

Presented to

The Faculty of the Department of Geography

San Jose State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Cammie d'Entremont Partelow

December 2008

SAN JOSÉ STATE UNIVERSITY

The Undersigned Thesis Committee Approves the Thesis Titled

USING GIS TO DEPICT RESOURCE RISK FROM PROBABLE *CANNABIS* CULTIVATION SITES

by
Cammie d'Entremont Partelow

APPROVED FOR THE DEPARTMENT OF GEOGRAPHY

12-Nov-2008

Dr. Richard Taketa, Associate Professor, Department of Geography Date

Dr. Gary Pereira, Assistant Professor, Department of Geography Date

11-13-08

Dr. M. Kathryn Davis, Assistant Professor, Department of Geography Date

11-6-2008

Dr. Leonard Gaydos, Date
Acting Regional Executive, United States Geological Survey

APPROVED FOR THE UNIVERSITY

12/2/08

Associate Dean Date

ABSTRACT

USING GIS TO DEPICT RESOURCE RISK FROM PROBABLE CANNABIS CULTIVATION SITES

by Cammie d'Entremont Partelow

Water diversion, fertilizer pollution, and the destruction of natural and cultural resources are among the impacts of illicit marijuana cultivation on public lands. Yosemite National Park and Sequoia/Kings Canyon National Park are challenged with mitigating this activity. Two habitat suitability model approaches, multiple logistic regression and weighted overlay, were compared in an effort to identify a best predictive model for marijuana grow site locations. The models analyze resource attributes and human activity. Resource criteria include attributes that are essential for plant growth such as slope, aspect, soil depth, and canopy. Human factors are site selection criteria that are unrelated to plant growth and include proximity to water and proximity to roads. This thesis presents an overview of the methods, data, results, lessons learned, sample output, and next steps. GIS modeling and analysis of this type is a valuable and efficient means to support resource protection and law enforcement on public lands.

ACKNOWLEDGEMENTS

Completion of this research and submission of this thesis is due in large part to the wonderful people that supported and encouraged me. Innumerable thanks are owed to Yosemite National Park, Chief Ranger, Steve Shackelton for his unwavering enthusiasm and sponsorship. Also, much appreciation is extended to the many other National Park Service rangers, investigators, and staff whose collaboration and feedback made this research possible. I would like to thank my advisor, Dr. Richard Taketa, and my committee members, Dr. Gary Pereira, Dr. M. Kathryn Davis, and Dr. Leonard Gaydos, for their time, effort, and contributions. A very special thank you is owed to Jim Matthews for his constant willingness to answer questions and provide guidance. To all my family, friends, and coworkers, I extend my gratitude for cheering me on. And, last but not least, I am eternally grateful to my husband Mike for all his patience, proofreading, and his firm belief that I would succeed.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

The Organic Act of 1916 charges the National Park Service to "conserve the scenery and the natural and historic objects and the wild life therein ... for the enjoyment of future generations" in "national parks, monuments, and reservations." With the increasing incidence of illegal marijuana cultivation on public land, this task is more difficult and more dangerous than ever. All public land agencies in the Western U.S., including national parks, are battling illegal *Cannabis sativa* (marijuana) growth within their boundaries. According to the Domestic Cannabis Cultivation Assessment produced by the National Drug Intelligence Center (2007), the number of plants eradicated from national forest lands in California nearly doubled from 2004 to 2006. Mexican drug trafficking organizations (DTO) are choosing to use public lands to grow marijuana rather than crossing tightened borders, and park rangers must modify their tactics to successfully protect resources against this threat (Coile, 2005; Whitehouse, 2005). Mexican nationals are hired by the DTOs to live in the area to tend and protect the plants. The damage caused by living in the area includes poaching of wildlife, excessive garbage, and human waste (U.S. Department of Justice, National Drug Intelligence Center, 2007). A recent cultivation site yielded "over 1,000 pounds of garbage" (Shilling, 2008). The understory is cut to make room for camps and marijuana plants, which allows non-native, invasive plants to take hold (Whitehouse, 2005). The use of park land to grow marijuana not only presents a risk to natural and cultural resources but also to the lives of park staff and visitors. Hikers are threatened by armed DTO guards, native plant habitats are destroyed, and ecosystems are polluted (Coile, 2005; Whitehouse, 2005; S. Shackelton, personal communication, June 26, 2007).

Identifying and then proactively monitoring sites at risk of impact from marijuana growers may facilitate the protection of resources before they are disturbed or harmed. Knowing the locations of and risks to resources allows park managers to better protect them (van Manen, Young, Thatcher, Cass, & Ulrey, 2005). Locating areas with planted marijuana (known as grow or cultivation sites) in national parks is a task for law enforcement experts who draw on experience to guide them. A geographic information system (GIS) which records information about spatial locations is an excellent means to capture and analyze such data. Anecdotal knowledge of various projects by individual park or forest divisions abounds, but a search for comprehensive information was mostly fruitless. Investigation into various local projects revealed that often the individual who created the model had changed jobs, the model was not being used, or time constraints did not permit further development or updates. Local agencies lack resources to facilitate continued development and adequate documentation. Research revealed abstracts for several modeling projects performed by various organizations, but again full articles or more in depth model information was difficult to locate. Three research projects conducted between the early 90s to the present were discovered.

An expert system is briefly discussed in the 1994 ACSM/ASPRS Annual Convention & Exposition conference proceedings. This system used expert knowledge to identify the factors which were slope, aspect, proximity to transportation network, proximity to water, distance from population centers, and forest cover (Fung & Welch, 1994). A location is selected by the user and the system indicates if the location is a possible cultivation site or not. Fung and Welch (1994) state "On the basis of the attribute information and the knowledge about Cannabis cultivation coded in the knowledge-base, the decision-support system determines the likelihood the selected location will be used as a growth site" but an algorithm is not explained. According to the evaluation, the results were promising with 83% of actual sites in predicted areas. Additional documentation was not found and attempted contact with the authors was not successful; therefore the continued use or current state of the system is unknown.

The Canadian Police Research Center (CPRC) researched a process that incorporated spatial analysis and image analysis in 2000 (Howell, 2002). The spatial analysis was a composite layer that was generated using the following predefined parameters: within 50 meters from a water source, within 500 meters from road access, within a cutblock or wetland area, under 1200 m elevation, south facing, and on government land (Howell, 2002). The composite layer was queried to select likely cultivation sites. The results were evaluated as good but only one site was verified and actual accuracy is not mentioned. In the conclusions to the study, the author expresses confidence in employing a GIS for exposing grow sites even with the limitations of the study (Howell, 2002). If this project was developed further is unknown.

The National Guard Bureau created a Decision Support System in conjunction with the Center for Higher Learning at the Stennis Space Center in Mississippi and Georgia Tech Resource Institute in Atlanta, with the same intention as this research:

prediction and planning. A 2004 reference to the system listed authors and developers who were supportive of this research and provided information. According to Jim Matthews (2007), one of the developers, the system is an Artificial Neural Network (ANN) that was trained with historical data and used variables including roads, streams, land cover, topography, rainfall, and first and last frost. A map that showed each point's similarity to known marijuana sites was generated for the study area. The ANN began with a univariate analysis of 100 parameters which was paired down to 20-25 (J. Matthews, personal communication, October 19, 2007). One downside to an ANN is that the decision process is internal. How the system arrives at an answer is hidden from the user. This system was not used in 2006 but was used by MS Bureau of Narcotics & MS National Guard previously and an attempt to use the system nationally was hampered by environmental differences and the varying levels of technology (J. Matthews, personal communication, October 16, 2007). A cluster analysis revealed that grow sites in the Central Valley had different qualities from grow sites in the rougher terrain of California. Similar analysis was performed for other states as well. Development and research was continuing on the project at the time of communication with Mr. Matthews.

These three projects illustrate that experts and scientists have attempted to use GIS as a tool to locate outdoor cultivation sites. GIS was successful in recognizing potential areas but concise, published documentation is lacking. Also highlighted is the lack of continued use or continued research and publication. Individual agencies probably tried GIS model projects which are likewise unpublished. Modeling areas of risk by incorporating spatial information and known factors of influence for site selection in a GIS will provide law enforcement an advantage in safeguarding resources by allowing more targeted surveillance.

1.2 Objective

The objective of this study is twofold. The first goal is to develop a scientifically based, defendable approach to model areas most likely at risk for illegal marijuana growing on federal land in the Sierra Nevada mountain range in California (Sierra Nevada). The second objective is to highlight areas of conflict between probable cultivation sites and known natural and cultural resource locations.

This research developed and compared two models using known data points: a multiple logistic regression model and a weighted overlay model. In the past, overlay models based solely on expert knowledge were used to show areas at risk. Two logistic regression models were developed based on different sample sizes and compared to identify which sample size would best represent the phenomena. The model approaches were selected for comparison based on a review of literature, current practice, and the data types used for the models. All models utilized publicly available data in conjunction with the coordinates for previous cultivation sites. Previous grow locations are monitored for additional activity and consequently the data are considered sensitive and not available for public use. The model approach which best represents areas most likely at risk for marijuana cultivation on federal land in the Sierra Nevada was used to visualize the risk to natural and cultural resources. Resource locations are maintained by the National Park Service (NPS) and were added as layers to the risk plots to show areas of concurrence.

2. Data and Methods

2.1 Study Area

The study area consists of two plots. A rectangular area surrounding and including Yosemite National Park (Yosemite) defined by the following UTM meter coordinates: 309731 4231097; 235067 414925. The second study area surrounds and includes Sequoia/Kings Canyon National Park (Sequoia) and is demarcated by the following UTM meter coordinates: 292579 3993847; 425929 4143277 (Figure 2.1).

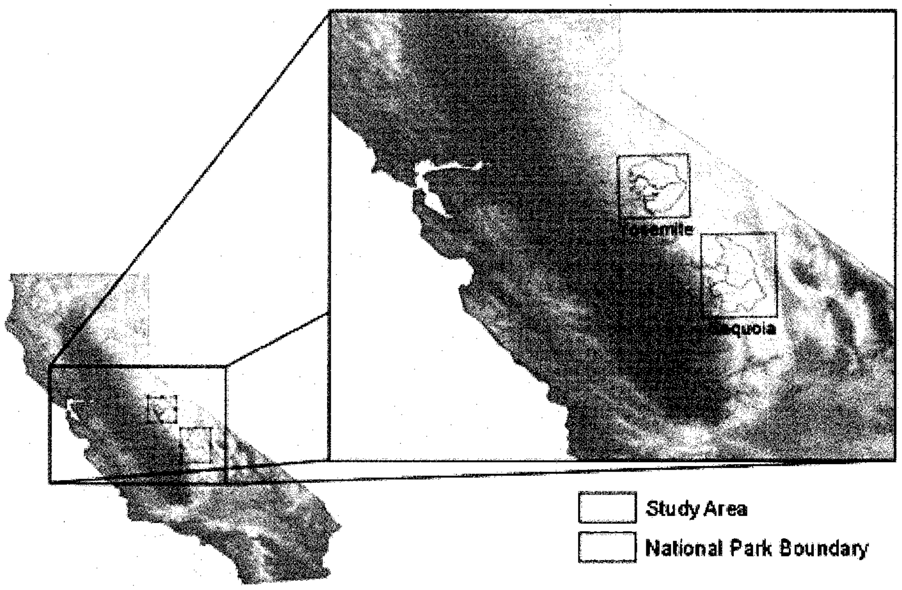

Figure 2.1: The study area is in the Sierra Nevada in California and includes Yosemite National Park and Sequoia/Kings Canyon National Park

Both parks have a diversity of threatened resources, a history of illegal *Cannabis* growing activity, are within relative proximity to each other, and have an experienced ranger staff committed to eliminating this activity. Although both parks in this study are in the Sierra Nevada range, the differences between them could prove too great to facilitate accurate, flexible model building by employing the data from only one. Yosemite is more north in the Sierra Nevada and Sequoia has higher elevations. Marijuana seizures increased from 10,000 marijuana plants eradicated from Yosemite in 2004 to nearly 21,000 in 2008. The problem is burgeoning in the Yosemite area and the data consists of only eighteen records. Sequoia has contended with this problem since the 1980s and has a larger database of marijuana cultivation locations.

2.2 Model Methods

A habitat model based on a discriminant analysis was used in Shenandoah National Park to successfully predict resource sites for endangered species at risk of poaching (van Manen et al, 2005). A discriminant analysis is a statistical model that groups events into two or more categories based on linear correlations between variables. In this type of model, data should be normally distributed. Normal distribution is a symmetrical probability distribution based on mean and variance (spread). Logistic regression analysis is a similar type of statistical analysis also applied to habitat modeling (Bonn & Schroder, 2001; Guisan & Zimmermann, 2000; Pearce & Ferrier, 2000). Logistic regression analyzes the independent or predictor variables to determine their correlation to the dependent variable (Hosmer & Lemeshow, 2000). This research is better suited to a logistic regression method considering the data utilized is a combination of qualitative and quantitative data and the outcome is a dichotomous result of presence or absence (Carroll, Zielinski, & Noss, 1999; Pereira & Itami, 1991). Logistic or binary regression yields the probability that an event will occur based on the values of known variables but it does not assume a normal distribution as discriminant analysis. Although statistical models are used frequently for predictive or descriptive modeling, land management agency law enforcement officers and rangers have historically utilized only overlay models to display areas at risk for cultivation.

As is often the case in habitat models, overlay models are based on expert knowledge to identify areas (Johnson & Gillingham, 2004). Habitat Suitability Models are used to identify plant and animal habitats (Gross, Kneeland, Reed, & Reich, 2002; Guisan & Zimmermann, 2000; Ortigosa, De Leo, & Gatto, 2000). Models are based on key factors or variables which relate to the species being modeled, and are selected through an understanding of the requirements for species success and the ability of the variable to be used in the model (U. S. Fish and Wildlife [USFW], 1981). Variables that signify potential marijuana cultivation sites were identified for inclusion in all models. A combination of expert knowledge and commonly used environmental variables was used to determine the parameters for the two models (Clevenger, Wierzchowski, Chruszcz, & Gunsun, 2002). Some of the uncertainty introduced by varying expert opinions (Johnson & Gillingham, 2004) may be reduced by using this combination. The experts for this study are the law enforcement rangers, investigators, and officers that locate and eradicate *Cannabis* grow sites. Variables were characterized into two groups: environmental and human.

Common plant or environmental requirement variables were selected based on literature reviews. The variables are elevation, slope, aspect, soil depth, water, vegetation and canopy cover (Gross et al, 2002; Guisan and Zimmerman, 2000; USFW, 1981; van

Manen et al, 2005). These factors affect plant viability. Although law enforcement experts stated that aspect was not a relevant variable based on recent cultivation activity, an analysis of the aspect of known grow sites in Sequoia revealed that a majority of known sites facing north did exist in 2006 but did not otherwise indicate a pattern of change. With the exceptions of 2002 and 2006, south and west directions were most prominent. As a result of this analysis, the variable was included in the development of the model. The variables remaining to be defined were selected to represent factors driven by human influence.

Social factors, variables not associated with plant biological needs but relevant to site selection, were recommended by law enforcement rangers who monitor sites for indications of activity. Proximity to water, trails and roads are important variables dominated by human choices. The knowledge of human impact which resides with the experts can guide the selection of objective data that might otherwise be overlooked. For example, proximity to roads is a consideration for the delivery of supplies (food, fertilizer, etc.) to the individuals that live at the grow site. Accordingly, a variable that represents a site's distance from a road is necessary. Other examples are environmental factors adapted to fit the needs of the marijuana growers including the redirection of water sources through tubing and the thinning of native plants. The individuals who grow marijuana at these locations are motivated by a goal to harvest the crop and avoid detection. They alter their behavior for unknown reasons. Eventually, a parameter that represents the mitigation activity of law enforcement rangers may need to be included. For now, each model considers the variables as defined above.

The logistic regression model employed the statistical package SPSS to perform the regression analysis on data derived from the cultivation and non-cultivation sites. The analysis generated a constant and a coefficient (ß coefficient) to represent the contribution of each variable to variations in the dependent variable which is location in this calculation. Canopy, land cover, and aspect are nominal data and were grouped into categories. Canopy and land cover were categorized in groups as defined by the data source (see appendix) with six and five categories respectively. Aspect data were grouped into nine categories: the four cardinal and four inter-cardinal directions, and flat. Dummy variables representing each group above and ß coefficients for those variables were created in the analysis (Pereira & Itami, 1991). This increased the variables and ß coefficients in the equation to 23 plus a constant. The ß coefficients were used in algorithms that assigned a value of presence/absence to the study areas and to categorize test points as cultivation site or non-cultivation site

The algorithm sums the products of variable values multiplied by the coefficient. This results in a value based on the following equation:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \ldots \beta_{22} x_{22} + \beta_{23} x_{23}$$

where β_0 is the constant, β_n is the coefficient, and x_n is the variable value. To obtain the likelihood estimation or probability, the z value is transformed using exp(z)/(1+exp(z)). The resulting value represents the probability that an area is or is not a cultivation site on a scale of zero to one with values closer to one less likely to be a grow site. Each transformed result is interpreted into a presence/absence result of cultivation site/non-cultivation site using the SPSS default classification cut-off of 0.5. All values equal to and less than 0.5 are potential grow locations. This cut-off appeared to provide good results for the models but the value can be changed.

The algorithm was applied to raster files to generate a plot of areas most likely to be cultivated using ESRI ArcGIS 9.2 (ArcGIS) Spatial Analyst Tools. Raster files are grids of pixels or cells each with a value and the values can be used in equations. For this analysis, a raster file with values for each variable was needed. The first step was to create raster files for the proximity and categorical variables. The Path Distance function in ArcGIS was used to generate raster files for each distance measure: proximity to water, trails, and roads. Path Distance considered elevation as part of the measure but an impedance value was not entered. The cell values represent the distance of the cell from the feature (stream, trail, or road). Next, the appropriate area was extracted for each dummy variable for canopy, land cover, and aspect using the Extract by Attribute function and saved as a new raster.

One issue encountered during the extraction process was that the extraction will only function on raster files with integer values. The aspect file cell type is a float which allows for decimals and higher precision. The conversion from float to integer using ArcMap conversion truncated rather than rounded the values. Therefore, values of 22.99 were saved as integer values of 22. About one half of a degree of precision was lost at each end of each range during the extraction step to create the dummy variables thereby compounding the error. In the final raster, it would be realistic to expect that cells at the beginning and end of each range were misclassified. To correct this problem, the floating point values were rounded and converted to integer using raster calculator prior to extraction. The following equation, based on an ESRI online technical article, was applied to the aspect raster dataset:

YoseAspectRnd = INT(CON([YOSEASPECT] > 0,CON(ABS([YOSEASPECT] - INT([YOSEASPECT])) >= 0.5, CEIL([YOSEASPECT]), FLOOR([YOSEASPECT])), CON(ABS([YOSEASPECT] - INT([YOSEASPECT])) >= 0.5, FLOOR([YOSEASPECT]), CEIL([YOSEASPECT]))))

The resultant raster file was in integer format and values could then be extracted into the correct dummy variable category.

Each dummy variable raster was reclassified to one for each cell with a value and zero for cells with no data. The raster files were multiplied by ß coefficient and combined into one raster whose cell values represented the ß values for each dummy variable. Recombination of dummy variables was performed using the mosaic function. The multiplication of each raster, the summation of all layers, and the log score calculation was performed in ArcGIS with the raster calculator. This process generated a final raster that visualized the probability of cultivation locations. The regression model deemed best by evaluation techniques discussed later in this text was plotted for Sequoia as described above and compared to the overlay model.

The overlay model was constructed using the minimum and maximum cell values of the sample points which were calculated in Microsoft Excel (Excel). The overlay model used only presence data to determine the risk area thus, to increase efficacy, all Sequoia actual cultivation sites were utilized. This process was used in an attempt to reduce expert bias and use actual site data to develop the model. Primary work region affects the figures that are presented by experts. The values were calculated for each variable and a range was established using the minimum and maximum values (Table 2.1). The evaluation of presence only data is an important difference from the logistic regression model which evaluates both presence and absence data to determine the ß coefficients.

Table 2.1 - Minimum and maximum values for variables based on the known cultivation sites

	Min	**Max**
Road Proximity	95.650	2876.340
Trail Proximity	39.481	5838.820
Stream Proximity	1.906	806.212
Canopy Cover	0.000	96.000
Land Cover	41.000	52.000
Soil Depth	43.000	117.000
Elevation	686.177	1817.527
Slope	3.815	108.983
Aspect	15.474	359.742

Each raster file was reclassified based on the range and cells with values between the minimum and maximum number receiving a value of one to represent risk and the remaining cells were given a value of zero to represent no risk. The raster files created for distance were reclassified for the proximity variables. A weighted overlay with equal weight was performed to create the area of anticipated risk. All zero values were assigned to no data in the scale and values of one were assigned a one.

2.3 Model Evaluation

The validation of models is a debated subject (Rykiel, 1996). Rykiel (1996) suggests defining the purpose and context of models prior to validation attempts. Guisan and Zimmerman (2000) advocate adopting the term "evaluated" in lieu of "validated." Another consideration when evaluating ecological presence/absence models is that an area in which the species is not observed does not mean the habitat is not suitable only currently unused. With these thoughts in mind, the models were evaluated to see if the predictive ability would be greater than chance (greater than 50%). SPSS provides several measures with the results to evaluate the process while the overlay method had only one.

Some of the several SPSS outputs for assessment or evaluation of a regression model are the Model Summary, the Omnibus Test of Model Coefficients, the Hosmer and Lemeshow test, and the classification tables. Applying the resultant ß coefficients to the test points and creating a classification table to compare the actual to the predicted observations is another means of evaluation (Pearce & Ferrier, 2000). The final evaluation technique consists of plotting the area and comparing where the known sites are to the area indicated as likely locations (Pereira & Itami, 1991). Consideration of the results inclusively is important because caveats exist for many of the tests. The SPSS evaluation outputs and application of the ß coefficients were used as a means to select which sample size generated the best regression model and which would be plotted and compared to the overlay model. Overlay models are evaluated based solely on the results of plotting.

To measure the effectiveness of plotting, the area deemed likely for cultivation sites was converted to polygons. The select function in ArcMap was used to select all points that were contained by the polygons. A second measure of proximity was performed using the same tool. The distance from the polygon was measured for points that were not contained by the polygons. The second measure was included as further understanding of the model results.

2.4 Data

The methods described above require data for training and for testing. Training is the process by which the model is developed. Known locations of grows and associated points were provided by Yosemite and Sequoia as point data. Known grow locations are the presence data and associated points – points associated with the grow location but which are not cultivation sites such as camps, dumps, or water lines – are considered absence data. The Yosemite data included 18 records, 12 known grow sites and 6 target sites. Target sites, areas where experts anticipate cultivation sites to be developed, were considered as non-cultivation sites for the test. The Sequoia dataset was accompanied by a caveat that coordinate data prior to 2000 was potentially faulty and had 17 records without coordinates recorded. Selecting all points with coordinates and a date after 12/31/1999 resulted in a subset of 131 records, 84 of which were known cultivation locations. The remaining records represent other points associated with the growing process but which are not cultivation sites such as camps, dumps, or water lines. The total dataset consisted of 96 known marijuana locations and 53 associated points.

To increase the total sample size and supplement the non-cultivation site subcategory, eighty random points were selected using the randbetween function in Excel and added to the dataset as non-cultivation sites: 60 from Sequoia/Kings Canyon study area and 20 from the Yosemite study area. These pseudo-absence points pose a potential risk because they are not confirmed as actual absence locations or they are closely associated with an actual cultivation location.

Two training datasets were randomly selected from the Sequoia points again utilizing the randbetween function in Excel. The first sample had 115 sample points (50 cultivation and 65 non-cultivation locations) and the second had 159 sample points (72 cultivation and 84 non-cultivation sites). The last dataset represents 75% of the total grow sites. This percentage representation of presence data was suggested by Pereira and Itami (1991) and van Manen et al. (2005). The number of non-cultivations sites is greater than the number of cultivation sites in both sample sets (van Manen et al., 2005). The remaining Sequoia points and all Yosemite data were reserved as test points.

The National Elevation Dataset 1 Arc Second (DEM) raster, National Land Cover Dataset 2001 – Land Cover, and National Land Cover Dataset 2001 – Canopy were downloaded from the National Map Seamless Data Server in ArcGrid format. The DEM, land cover, and canopy are raster files with a spatial resolution of 30 meters. Slope and aspect are derived from the DEM. The road and trail data are vector files and were downloaded from the National Park Service (NPS) Data Store. Stream data for Sequoia was downloaded from the NPS Data Store and for Yosemite is a combination of vector files received from Yosemite's GIS Office and the National Map Seamless Data Server.

The National Map Seamless Data Server file provided information for the eastern and southeastern portion of the study area for which the GIS Office did not provide stream information. All raster files were converted to a cell size or spatial resolution of 30 meters.

The Soil Survey Spatial and Tabular Data (SSURGO 2.2) soil dataset available for Yosemite is unavailable for Sequoia. State Soil Geographic (STATSGO) is a lower resolution dataset but is available for both parks. STATSGO provides a range for depth-to-bedrock data. The CONUS-Soil dataset was developed by the Center for Environmental Informatics at The Pennsylvania State University (http://www.soilinfo.psu.edu/) for modeling. The dataset uses STATSGO and provides a mean depth-to-bedrock value (Miller, 1998) but is only valid to a depth of 152 cm (60 in). Values of 152 could represent a greater depth. The data was downloaded from CONUS-Soil and used for both study areas to maintain consistency.

North American Datum 1983 (NAD 83) UTM zone 11 is the standard projection for Yosemite and Sequoia. Files were reprojected as necessary and the North American Datum Conversion (NADCON) transformation was used files originally projected with North American Datum 1927 (NAD 27). Known grow site locations were received in degrees decimal minutes in a NAD 83 projection for Yosemite and in UTM zone 11 meters from Sequoia. Yosemite points were converted to UTM meters.

2.5 Model Discussion and Results

Statistical tests require the selection of a critical value or, restated, the level at which it is acceptable to be wrong. For all test, the critical value was set at 0.05. This means that the results are expected to be correct 95% of the time. Statistics with significance levels equal to or smaller than 0.05 are considered statistically significant (SPSS Topics Logistic Regression).

An assessment of the results from SPSS revealed that the model based 115 point sample performed well as a model. The omnibus test of model coefficients is a measure of the significance of the addition or subtraction of a variable or variables. The significance denotes the probability of obtaining the chi-square value without the independent variables (SPSS Topics Logistic Regression). The significance value was less than 0.05 indicating that the addition of the variable was statistically significant (Table 2.2).

Table 2.2 - SPSS 115 point model Omnibus Test of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	91.442	23	.000
	Block	91.442	23	.000
	Model	91.442	23	.000

R^2 represents the variation explained by the model in linear regression (Menard, 2002). A true R^2 cannot be computed for Logistic regression therefore pseudo R^2 values are calculated using the -2 Log likelihood. A pseudo R^2 value closer to 1 indicates that more of the variation is explained by the model. As indicated by Table 2.3, the Nagelkerke R Square indicates that 74% of the variation is explained by the model.

Table 2.3 – SPSS Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	66.020[a]	.548	.736

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

One note of caution Menard expresses is that if the sample size is too small a substantively significant R^2 may still be caused by random variation. The sample size used in this study is small thus this measure was considered in conjunction with other results for model evaluation.

The next measure is the Hosmer and Lemeshow test, a goodness-of-fit test of the null hypothesis that the model does not adequately fit the data. According to Garson (2008) this measure is "more robust than the traditional chi-square test, particularly if continuous covariates are in the model or sample size is small." The cases are divided into deciles by predicted probabilities and a chi-square statistic is computed from observed and expected frequencies to obtain the measure for this test (Menard, 2002). In this test and contrary to other chi-square evaluations, an insignificant chi-square (greater than 0.05) means the model is a good fit. The chi-square and degrees of freedom value are similar for this model suggesting that the null hypothesis is true but the significance was greater than 0.05 demonstrating that the model is a good fit (Table 2.4).

Table 2.4 - SPSS Hosmer and Lemeshow test results for the 115 point sample logistic regression analysis

Step	Chi-square	df	Sig.
1	9.528	8	.300

Reviewing the SPSS results for the 159 point model shows that the addition of the variables was a significant change (Table 2.7) and that the model has about 68% of the variation explained (Table 2.8).

Table 2.7 - SPSS 159 point model Omnibus Test of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	112.154	23	.000
	Block	112.154	23	.000
	Model	112.154	23	.000

Table 2.8 - SPSS 159 Point Model Summary from SPSS

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	106.850[a]	.506	.677

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

In the Hosmer and Lemeshow test for the 159 point model, the chi-square and degrees of freedom were not similar signifying that the hypothesis is valid and the significance was quite high (Table 2.9) indicating that the model is a good fit.

Table 2.9 - SPSS Hosmer and Lemeshow test results for the 159 point sample logistic regression analysis

Step	Chi-square	Df	Sig.
1	2.459	8	.964

The classification table showed an overall correct percentage slightly lower than the 115 point model but still acceptable at 84.3% (Table 2.10) and the percentage correct for cultivated sites was slightly higher. This was also an increase over the baseline results of 54.7% (Table 2.11).

Classification tables reflect the predictive abilities of the model. SPSS runs a preliminary regression without any variables as a baseline before processing the variables for the logistic regression (Table 2.5). This table shows the percentage correct if all points were predicted to be non-cultivated sites: 56.5% or slightly better than chance. Consideration should be given to the difference between the two classification tables as well as the overall percentage of correct predictions (Garson, 2008). The 115 point model had a high percentage correct, overall 86.1%, in the classification table from SPSS (Table 2.6) and was overall much better than the constant only model.

Table 2.5 - SPSS generated baseline Classification for the 115 point sample without variables

	Predicted		
Observed	Cultivated	Non-cultivated	Percentage Correct
Cultivated	0	50	0
Non-cultivated	0	65	100.0
Overall Percentage			56.5

Table 2.6 - SPSS generated classification table for the 115 point sample logistic regression analysis

	Predicted		
Observed	Cultivated	Non-cultivated	Percentage Correct
Cultivated	44	6	88.0
Non-cultivated	10	55	84.6
Overall Percentage			86.1

Three of the four evaluation statistics for this model suggested that the model fit the data. The only statistic that reflected poorly on the model was the Chi-square value of the Hosmer and Lemeshow test. This logistic regression model is adequate considering all of the measures from the SPSS regression output.

Table 2.10 - SPSS Classification table for the 159 point sample logistic regression analysis

	Predicted		
Observed	Cultivated	Non-cultivated	Percentage Correct
Cultivated	66	6	91.7
Non-cultivated	19	68	78.2
Overall Percentage			84.3

Table 2.11 - SPSS Baseline classification table for the 159 point sample without variables

	Predicted		
Observed	Cultivated	Non-cultivated	Percentage Correct
Cultivated	0	72	0
Non-cultivated	0	87	100.0
Overall Percentage			54.7

All four evaluation statistics for the 159 point model indicated that the model was a good fit for the data. This logistic regression model is adequate considering all of the measures from the SPSS regression output. The next evaluation method does not occur in SPSS.

A manual application of the ß coefficients to the test points to verify if the regression coefficients can accurately categorize points is performed using Excel. The algorithm explained in the Model Methods section was applied using the values for each variable and the ß coefficients. A classification table was constructed to compare the predicted and observed values. This process produced an overall percentage correct classification lower than the classification table for the training set. In both cases, the results are significantly better than chance (50%). 70% of actual cultivation locations were correctly classified when the ß coefficients derived from the 115 point model were applied. The calculation incorrectly categorized 10 previous grow sites and 18 non-cultivation sites (Table 2.12).

Table 2.12 - Classification table for the 115 point regression test points analysis

Observed	Predicted		
	Cultivated	Non-cultivated	% Correct
Cultivated	23	10	70%
Non-cultivated	18	55	75%
Total	41	65	74%

The results from the 159 point algorithm calculation were lower overall but better in cultivation site discrimination than the 115 point equation. The number of correctly classified non-cultivation sites was the same with both equations but resulted in a lower percentage (64%) correct for the 159 point equation. 100% of previous grow location test points were classified as cultivated (Table 2.13).

Table 2.13 - Classification table for the 159 point regression test point analysis

Observed	Predicted		
	Cultivated	Non-cultivated	% Correct
Cultivated	12	0	100%
Non-cultivated	18	32	64%
Total	30	32	71%

When applied to the Yosemite test points, the results were again slightly lower at 72% and 71% overall but the pattern of correct classifications remained the same (Table 2.14 & Table 2.15). In both cases, a higher correctness of presence classifications, known as sensitivity, existed while correct absence classification was lower (Pearce & Ferrier, 2000). The 159 point model correctly classified ten, or 83%, of the known grow site locations while the 115 point sample had nine, or 75%, correct. The small number of test points causes the percentage correct to be vastly different for small changes in number.

Table 2.14 - Contingency table of Yosemite test points using the 115 sample point coefficients

Observed	Predicted			
	Cultivated	Non-cultivated	Total	% Correct
Cultivated	9	3	12	75%
Non-cultivated	8	18	26	69%
Total	17	21	38	72%

Table 2.15 - Contingency table of Yosemite test points using the 159 point coefficients

Observed	Predicted			
	Cultivated	Non-cultivated	Total	% Correct
Cultivated	10	2	12	83%
Non-cultivated	9	17	26	65%
Total	19	19	38	71%

Fielding and Bell (1997) and Pearce et al (2000) state that the accuracy measurement, or total percent correct, may be misleading. In both classification tables, the non-cultivation sites had a higher percentage of misclassifications. Further investigation revealed that the majority of misclassified non-cultivation sites were points associated with grow sites. 14 points or 78% of the misclassified non-cultivation sites in the 115 point sample and 17 points or 94% of the 159 point sample were near a known cultivation site. The target points in the Yosemite data represent 67% of the misclassification in the 159 point non-cultivation category. Errors which include areas in a presence result but are not observed, also known as Errors of Commission, on spatially correlated cells are known to be problematic (Fielding, 1997). This error is acceptable because points so closely related to a known cultivation location could be potential grow sites. As mentioned previously, discerning errors can be difficult because an area as yet unused may be used in the future.

The totality of evaluation statistics suggested that the two regression models were not extremely different but that the 159 point model was more sensitive and slightly more able to discriminate between cultivation and non-cultivation sites (Pearce, 2000). The 115 point model was set aside in favor of the model that used the 159 sample points. The final regression model evaluation was to apply the regression coefficients to the study area raster files to visualize the area likely to have grow sites.

When plotted in ArcMap, the anticipated cultivation area is in the southwest region of the study area. This area coincides with the locations of known grow sites and the area indicated by experts. The actual sites were displayed as points and overlaid on the plot (Figure 2.2). Four of the 84 training points are not contained within the area indicated as likely grow sites. All of the four plots are within 30 meters of an anticipated area and are adjacent to a point that is contained in the anticipated area. Plotting yielded a higher percentage correct than the applying the ß coefficients to the training or test points with 95% of known cultivation sites within the predicted area.

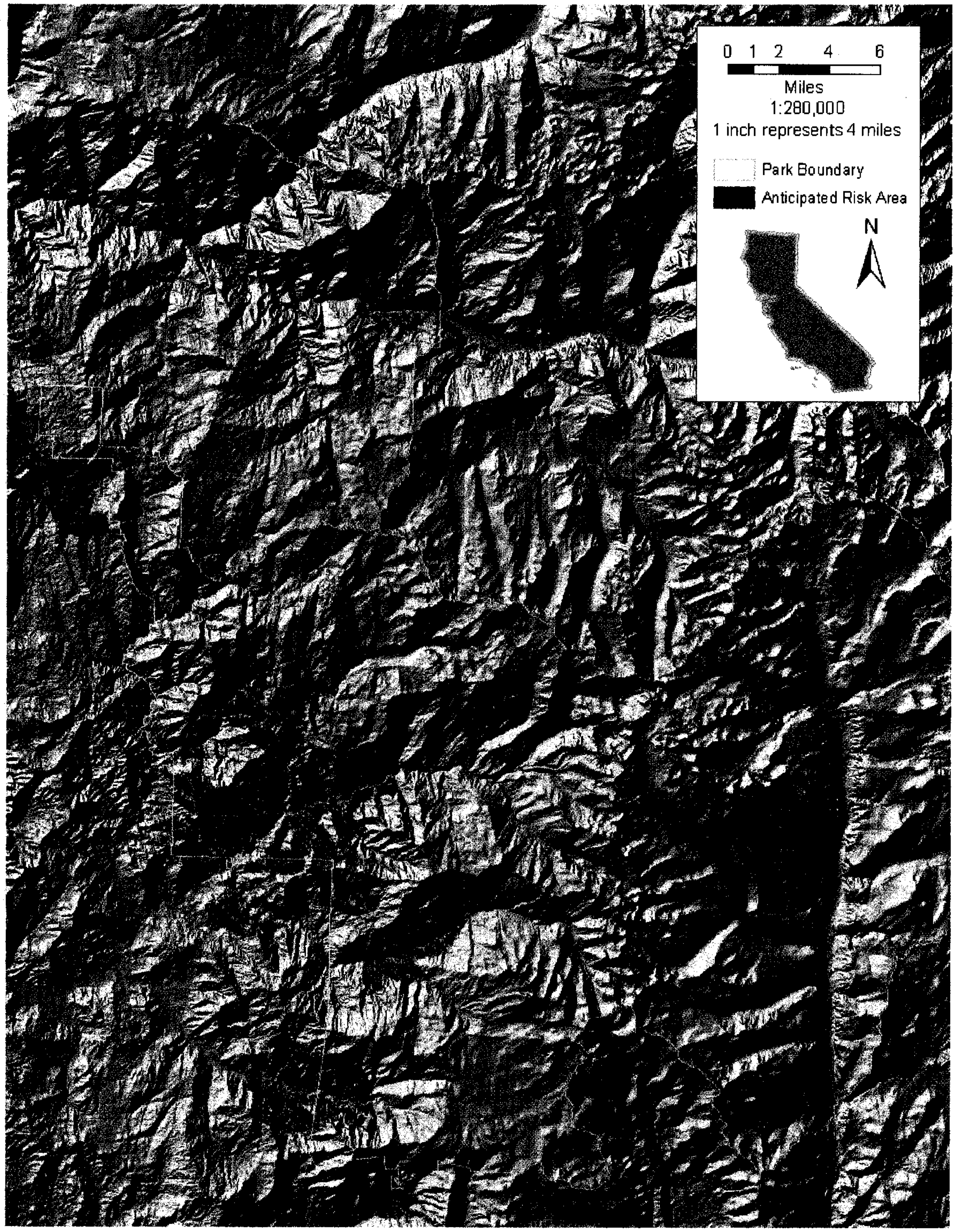

Figure 2.2: Sequoia study area weighted overlay model

The Weighted Overlay model based on all 84 actual cultivation points had 36 points fall outside of the area indicated as most likely. 57% of the known locations are in the area forecasted as likely growing locations. Only 14 points are greater than 100 meters away from a predicted presence cell. Even considering the proximity, the weighed overlay model remains less accurate than the logistic regression model.

A visual comparison of the two Sequoia plots (Figure 2.3) revealed that the overlay anticipates less total area for cultivation sites and the large predicted areas are disconnected from each other. The same general regions were highlighted by the regression and overlay models. Both models have anticipated sections that are small, one or two pixels, and predicted areas have small omitted portions. The overlay plot appeared to have fewer gaps with smoother edges. The logistic regression plot had areas that spread together and covered more of the region but the edges appear pixilated. The overlay plot did not highlight the canyons, a geographic feature that experts suggest is important, as well as the logistic regression plot. The overlay model is simpler in its development but it is less accurate.

When the results for the regression model were plotted and known sites overlaid for the Yosemite study area, the plot encompassed 14 of the 18 points (Figure 2.4). This result was better than the test point evaluation with 78% of the points falling in the predicted area. Two actual cultivation and two target sites were not contained in the plot area. All four points were within 30 meters of the area identified to have likely cultivation sites.

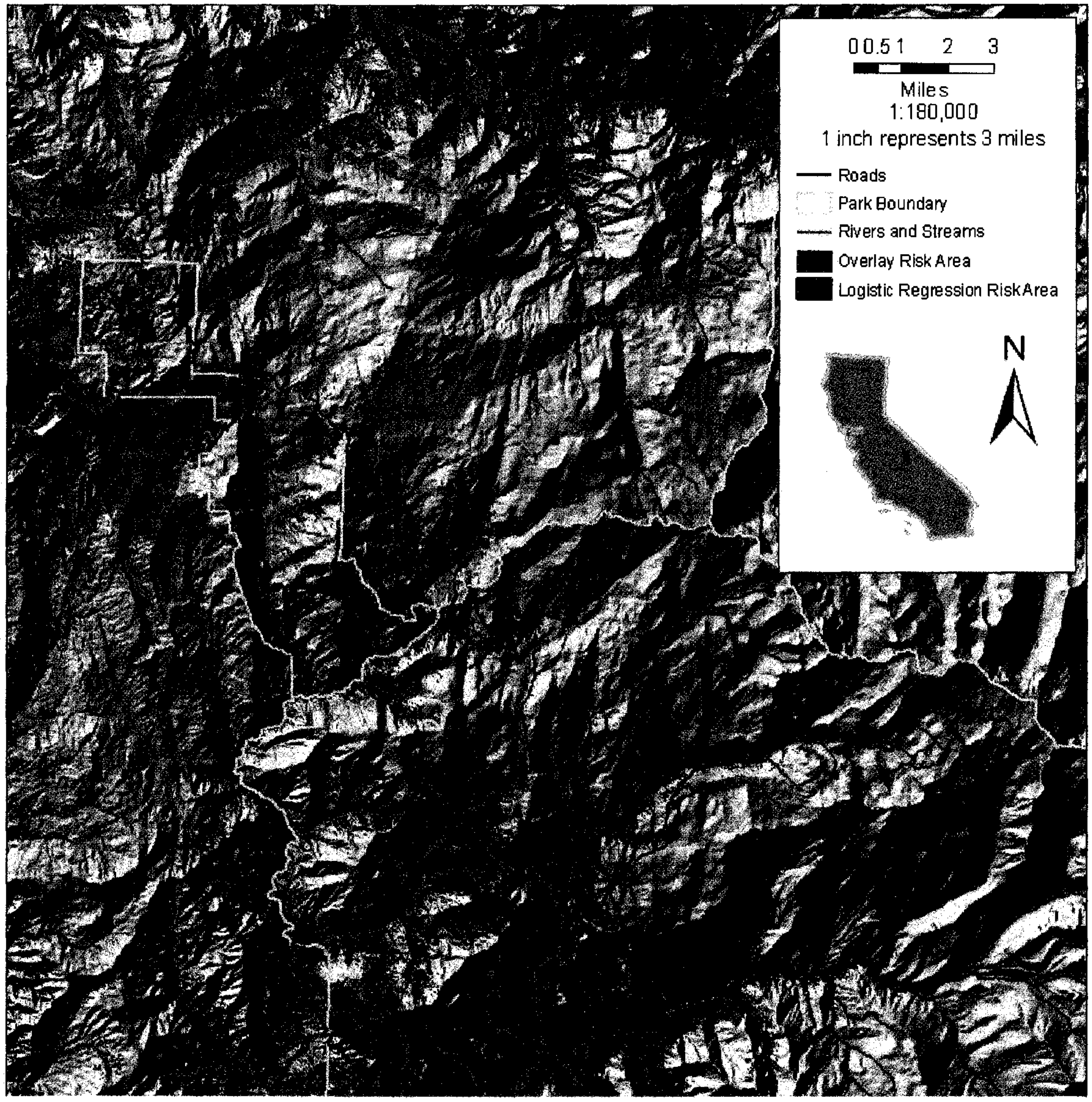

Figure 2.3: Close up of Sequoia overlay and logistic regression areas of risk overlaid for comparison

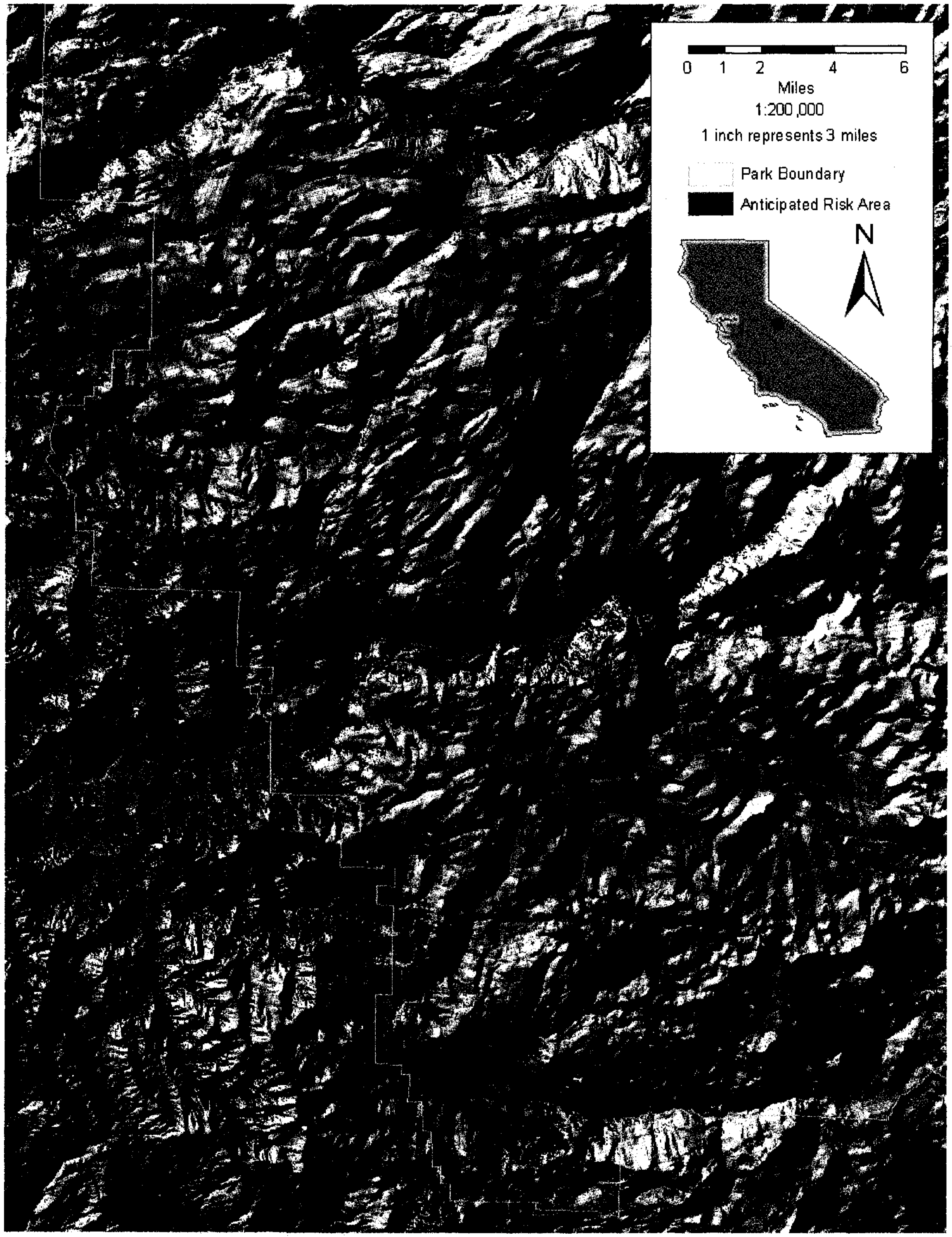

Figure 2.4: Yosemite study area regression model plot

2.6 Further Discussion

The data for this study was collected in February of 2008. In August 2008, one of the NPS investigators stated that one target site was eliminated as a suspect location and another was identified for surveillance. The coordinates of the removed site did not match any coordinates previously supplied. Both locations were identified as target sites in May and neither was part of the analysis data. When plotted, the new sites raised questions and a subsequent conversation revealed two key points that need addressing. Each will be discussed separately.

One new target was plotted and appeared in water. This point was approximately 170 ft from an area categorized as a likely cultivation area. This site is across the water from an eradicated site. Both sites were known to be cultivated by a local individual, not a Mexican DTO. These points do not correspond to the apparent DTO prerequisites for site selection. Local growers do not have the same modus operandi. The cultivation sites tend to be smaller, unguarded, and in some cases including this one, far from roads. A second analysis using only known DTO grows may more accurately reflect the expected location of larger size sites while the current model shows all possibilities. Both models could be used to locate grow sites. Data from additional sources will be required because Yosemite and Sequoia do not have enough representative points. Only six of the Sequoia sites are labeled as DTO managed. It is possible that points are incorrectly coded in the database.

Errors or inconsistencies in data recording were the second concern. The coordinates for another target site which was adjacent to a probable area were identified as possibly inaccurate. The point was selected using a map and a general idea of the location versus a global positioning system device (GPS) at the actual location. Although GPS units can be incorrect up to 20 feet, the coordinates garnered would be more accurate than the current technique which provides coordinates that are somewhat arbitrary. The investigator stated that the target may actually be in the area indicated by the model. This issue highlights the need for a GIS specialist. The NPS Pacific West Region would benefit from a dedicated GIS manager who could collect, maintain, analyze and report on regional data, and fulfill need for improvement and maintenance of models in hopes of continued and shared use as was mentioned in the introduction. Better collection and standardized data formats in a central repository would improve the data analysis. Individual parks have GIS personnel with varying skill levels and numerous responsibilities. A regional level person dedicated to working with investigators and rangers to eliminate cultivation sites would develop expertise that would benefit all parks in the region. Separate parks would not need to impose additional tasks on their GIS staff but could leverage the global skills and knowledge as is currently the practice with investigators. Such an individual could liaise with other agencies facing this same issue. Many national parks, including Yosemite, are surrounded by national forest or state or county public lands. Each of these agencies is faced with eradicating cultivation sites. Data from other agencies could be incorporated to improve the model and, again, benefit all.

2.7 Issues

The lack of a variable that represents the influence of human presence was most evident in the Yosemite plot. Yosemite Valley is highlighted as a likely area for cultivation. Although the environment is suitable for growing marijuana, the high level of human activity makes it less likely to be selected as a cultivation site. With that said, at least one small cultivation site was anecdotally noted in Yosemite Valley; so a variable that would represent level of activity was not investigated. A surrogate measure to characterize locations where population or development inversely influences location selection should be explored for possible inclusion in the equation.

Overall size of the probable growing area may be a factor in location selection. Grow site size is not currently measured but DTO growers are planting in locations which allow for large crops. In some cases, multiple cultivation sites are planted in proximity to each other. An estimate of preferred area could be determined based on an average of the number of plants eradicated and an estimated per plant area requirement. For example, if each plant requires 3 to 6 square meters growing space and the average number of plants eradicated per site is 1,000, then a minimum size could be 3,000 square meters. Locations below the minimum size could be eliminated from the risk area. This variable would eliminate some of the smaller disconnected sections from the predicted area.

3. Areas of Conflict

3.1 Identifying Conflict

A key goal for this process was to identify areas of potential conflict to resources. Resources that require protection are rare plants, cultural sites, and wildlife. Some species, such as the mountain yellow-legged frog, are in decline or threatened. Others, such as the Yosemite Orchid, are known to exist only in this region. Cultural sites, such as Native American or cavalry use sites, have historical value that is irreplaceable. To reach the goal of detecting risks to resources, rare plant locations, archeological sites, and some wildlife habitat locations were overlaid on the selected regression model. Yosemite maintains GIS layers representing natural and cultural resource locations.

The rare plant layer consists of over 100 species and includes ten plants that are listed as species of concern or rare by the federal or state government including the Yosemite Wooly Sunflower and the Yosemite Onion. The Mountain Lady Slipper, several sedges, and some sunflower species are considered rare in Yosemite but not in the federal or state protection lists.

Cultural resource sites are located all over. The Yosemite Valley Historic District is on the national register of historic places and includes many national historic landmarks such as the Awahnee Hotel. The archeological layer has sites such as caves with petroglyphs, Native American living areas, U.S. Calvary camps, gold mine locations and national historic landmarks.

Unfortunately, not all wildlife layers were provided for this research because some wildlife information is considered too sensitive to be released. Locations of

Spotted and Great Gray Owl nests, Flying Squirrel and Yosemite Toad habitat are included in the wildlife layer. Eventually a complete wildlife layer that includes ranges should be included to show the potential risk to all wildlife especially considering that growers are known to poach. Adding these layers to the likely risk plot showed several potential coincident areas and highlighted the risk to resources (Figure 3.1).

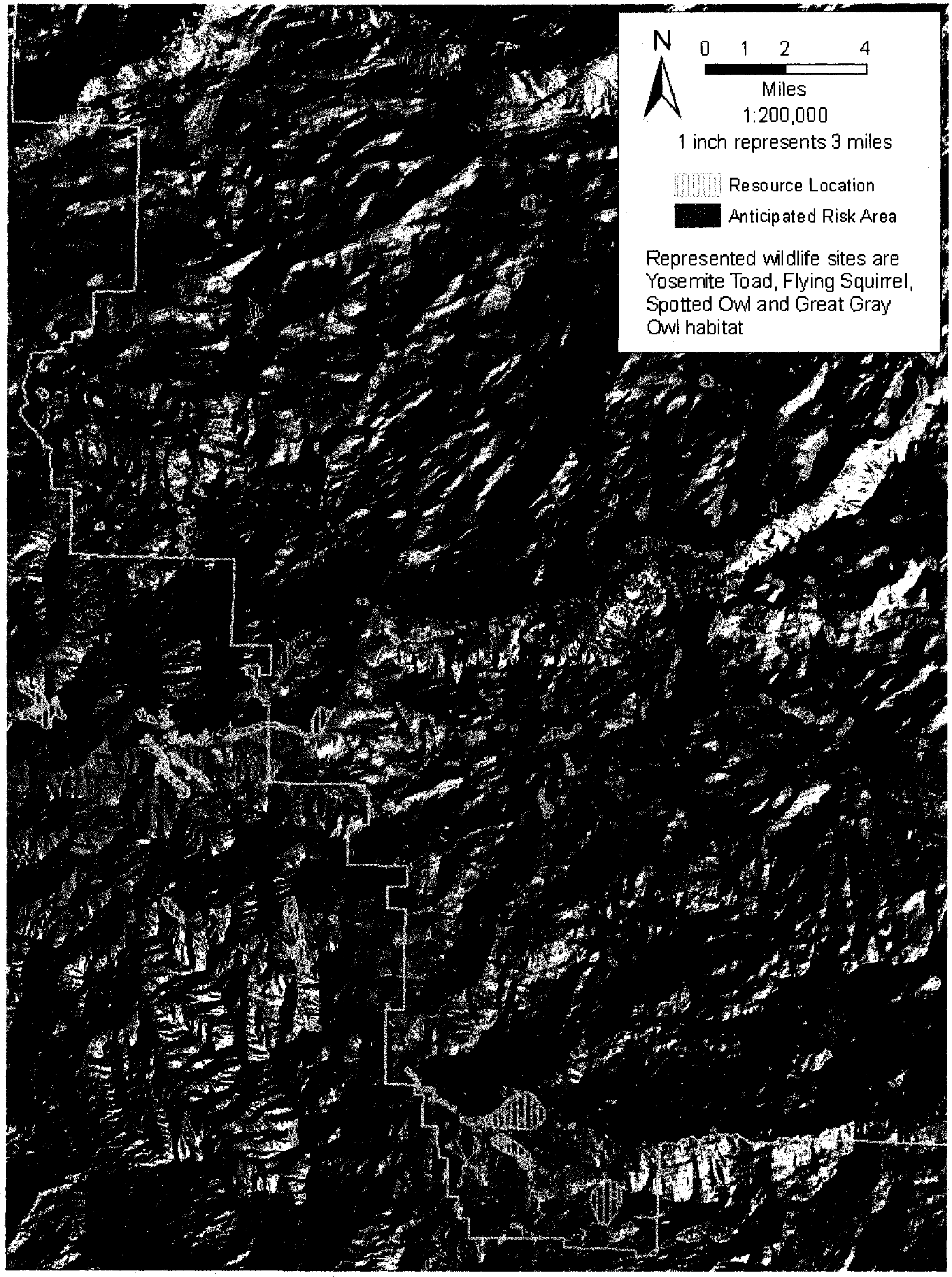

Figure 3.1: Yosemite area of conflict

Many resource locations overlap with the anticipated risk area. In fact, some resource locations may already be compromised. A close review of the plotting reveals three eradicated sites, two pictured below in Figure 3.2, in close proximity to archeological sites and rare plant habitat and one target site that coincides with a wildlife location.

Figure 3.2: Zoom of area of conflict highlighting two eradicated cultivation sites in proximity to resource locations

3.2 Discussion and Conclusions

A comparison of the models indicates that a model based on logistic regression analysis is more effective than a weighted overlay model at identifying areas most likely at risk for marijuana cultivation on federal land in the Sierra Nevada of California. Knowing that presence/absence models are difficult, if not impossible, to validate, the regression model has four evaluation methods that provide an ability to stand up to scrutiny while the overlay model has only one. Although the overlay model is easier to generate and can be completed completely in ArcGIS, the regression method is more scientifically defendable. Of the two regression models, the 159 point model is more sensitive with 100% of known grow sites in the test sample correctly classified. Therefore, the method that provides the best model for the purpose of this study is the logistic regression model that utilized 75% of available presence data and absence data to generate the ß coefficients.

The risk to protected resources is high. The plotting of the probable area for marijuana cultivation and resource locations underscores the risk. Illegal marijuana cultivation is a threat that must be addressed before resources are permanently damaged. Modeling likely cultivation sites using a GIS is a good way to visualize the potential risk to resources. This can guide resource managers and law enforcement rangers in their protection efforts as well as help illustrate the problem to policy makers. Using this process to manage areas of conflict may help reduce the habitat destruction and pollution that occurs as a result of marijuana cultivation activities by making an area less appealing to DTOs.

3.3 Next Steps

Ideas for future research beyond this project are plentiful. As an initial step, the items mentioned in the issues sections should be addressed. Establishing a variable to represent the influence of human activity and eliminating areas that are too small to be used for cultivation will create a more precise model. Incorporating the 2008 grow location data points and cultivation status updates will provide additional data to improve the model. This step should include both plotting the 2008 grow location points on the likely area plot, and recalculating the logistic regression equation using a new subset of 75% of the data. Incorporating data from neighboring agencies would enhance the model, and exploring regression models using only DTO grows, as discussed in section 2.6, may help better distinguish areas. Continuing analysis of variables and new data will improve the predictive accuracy of the tool.

Experimenting with impedance values in the path distance measure may improve the distance variable. Slope, vegetation type, canopy, and combinations of variables are possible candidates from the current variable set to represent impedance. The steepness of a slope or the thickness of the understory, which increase the difficulty in reaching a location, may affect site selection.

A conflict area that should be explored is the pollution of resources such as water and soil. Water and soil are both polluted by fertilizer and pesticides, which are not removed with plant eradication. These resources are not included in the current conflict identification nor are the implications of such tainted resources. A method to define, measure, and model the total area affected by cultivation sites should be determined to assess the total resource impact.

Finally, creating a tool that can be run in ArcMap by a lay user is important to facilitate use by rangers. This is especially important as long as individual parks are expected to manage this process. To be successful, the tool must not require a significant amount of time and energy to learn but should provide accurate visualization which will allow rangers to focus on preventing the establishment of cultivation sites. The projects that have not continued illustrate the need for a tool that can be used by a person that is not a GIS expert. By utilizing Python scripts, a tool that prompts users for the necessary data and does not require users to have a thorough understanding of ArcMap can be developed and initiated.

References

Bonn, A., & Schroder, B. (2001). Habitat Models and Their Transfer for Single and Multi Species Groups: A Case Study of Carabids in an Aluubial Forest. *Ecography, Volume 24, Issue 4, August 2001*, 483-496.

Campaign Against Marijuana Planting. (2007). Retrieved October 9, 2008, from Office of the Attorney General: http://ag.ca.gov/bne/camp.php

Carroll, C., Zielinski, W. J., & Noss, R. F. (1999, December). Using Presence-Absence Data to Build and Test Spatial Habitat Models for the Fisher in the Klamath Region, U.S.A. *Conservation Biology, Volume 13, No. 6*, pp. 1344-1359.

Clevenger, A. P., Wierzchowski, J., Chruszcz, B., & Gunsun, K. (2002). GIS-generated, Expert-based Models for Identifying Wildlife Habitat Linkages and Planning Mitigation Passages. *Conservation Biology, Volume 16 (2) April 2002*, 503-514.

Coile, Z. (2005). National Parks' Pot Farms Blamed on Cartels. *San Francisco Chronicle Nov. 18, 2005 Newsroom Yosemite Association 25 June 2007*.

Fielding, A. H., & Bell, J. F. (1997). A Review of Methods for the Assessment of Prediction Errors in Conservation Presence/Absence Models. *Environmental Conservation 1997, Volume 24*, 38-49.

Fung, D. S., & Welch, R. (1994). MODELING CANNABIS CULTIVATION IN NORTH GEORGIA. *ACSM/ASPRS Annual Convention & Exposition* (pp. 217-220). Reno: http://libraries.maine.edu/Spatial/gisweb/spatdb/acsm/ac94024.html, Accessed October 6, 2007.

Garson, G. D. (2008, May 21). *Logistic Regression.* Retrieved October 2, 2008, from Statnotes: Topics in Multivariate Analysis: http://faculty.chass.ncsu.edu/garson/PA765/logistic.htm

Gross, J. E., Kneeland, M. C., Reed, D. F., & Reich, R. M. (2002). GIS-based Habitat Models for Mountain Goats. *Journal of Mammalogy, Volume 83, Issue 1, February 2002*, 218-228.

Guisan, A., & Zimmermann, N. E. (2000). Predictive Habitat Distribution Models in Ecology. *Ecological Modelling, Volume 135, Issues 2-3, 5 December 2000*, 147-186.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression, Second Edition.* Hoboken: John Wiley & Sons.

Howell, C. T. (2002). *TR-03-01 Operation SABOT and Illicit Crop Information Management Using Satellite Imagery.* Richmond, British Columbia.

Johnson, C. J., & Gillingham, M. P. (2004). Mapping uncertainty: sensitivity of wildlife habitat ratings. *Journal of Applied Ecology, Volume 41* , 1032-1041.

Menard, S. (2002). *Applied Logistic Regression Analysis, Second Edition.* Thousand Oaks: Sage Publications.

Miller, D. A., & White, R. A. (1998). *A Conterminous United States Multi-Layer Soil Characteristic Data Set for Regional Climate and Hydrology Modeling.* Retrieved April 17, 2008, from Earth Interactions, 2: http://EarthInteractions.org

Ortigosa, G. R., De Leo, G. A., & Gatto, M. (2000). VVF: Integrating Modelling and GIS in a Software Tool for Habitat Suitability Assessment. *Environmental Modelling and Software, Volume 15, Issue 1, January 2000* , 1-12.

Pearce, J., & Ferrier, S. (2000). Evaluating the Predictive Performance of Habitat Models Developed Using Logistic Regression. *Ecological Modelling, Volume 133, Issue 3, 3 September 2000* , 225-245.

Pereira, J. C., & Itami, R. M. (1991). GIS-based Habitat Modeling Using Logistic Multiple Regression: A Study of the Mt. Graham Red Squirrel. *Photogrammetric Engineering and Remote Sensing, Volume 57, Number 11, 1991* , 1475-1486.

Rykiel, J. E. (1996). Testing ecological models: the meaning of validation. *Ecological Modelling 90* , 229-244.

Shilling, K. (2008, August 18). *North Cascades National Park Service Complex.* Retrieved October 7, 2008, from National Park Service: http://www.nps.gov/noca/parknews/major-marijuana-plantation-removed-from-ross-lake-national-recreation-area.htm

SPSS Topics Logistic Regression. (n.d.). Retrieved October 1, 2008, from UCLA: Academic Technology Services, Statistical Consulting Group: http://www.ats.ucla.edu/stat/spss/topics/logistic_regression.htm

U.S. Department of Justice, National Drug Intelligence Center. (2007). *Domestic Cannabis Cultivation Assessment 2007.*

U.S. Fish and Wildlife. (1981). *Standards for the Development of Habitat Suitability Index Models, Ecological Services Manual.* Accessed from http://www.fws.gov/policy/ESM103.pdf on February 26, 2008.

van Manen, F. T., Young, J. A., Thatcher, C. A., Cass, W. B., & Ulrey, C. (2005). Habitat Models to Assist Plant Protection Efforts in Shenandoa National Park, Virginia, USA. *Natural Areas Journal, Volume 25 (4), 2005* , 339-350.

Whitehouse, L. (2005, November 17). *NPCA | Testimony of Laura Whitehouse.* Retrieved October 7, 2008, from National Parks Conservation Association: http://www.npca.org/media_center/testimonies/testimony111705.html

Appendix: Categorical Variables Key

Land Cover

Code	Description	Class Code	Class Name
11	Open Water	1	Water
12	Perennial Ice/Snow	1	Water
21	Developed, Open Space	2	Developed
22	Developed, Low Intensity	2	Developed
23	Developed, Medium Intensity	2	Developed
24	Developed, High Intensity	2	Developed
31	Barren Land (Rock/Sand/Clay)	3	Barren
32	Unconsolidated Shore*	3	Barren
41	Deciduous Forest	4	Vegetated; Natural Forested Upland
42	Evergreen Forest	4	Vegetated; Natural Forested Upland
43	Mixed Forest	4	Vegetated; Natural Forested Upland
51	Dwarf Scrub	4	Vegetated; Natural Shrubland
52	Shrub/Scrub	4	Vegetated; Natural Shrubland
71	Grassland/Herbaceous	5	Herbaceous Upland Natural/Seminatural Veg
72	Sedge/Herbaceous	5	Herbaceous Upland Natural/Seminatural Veg
73	Lichens	5	Herbaceous Upland Natural/Seminatural Veg
74	Moss	5	Herbaceous Upland Natural/Seminatural Veg
81	Pasture/Hay	5	Herbaceous Planted/Cultivated
82	Cultivated Crops	5	Herbaceous Planted/Cultivated
90	Woody Wetlands	6	Wetlands
91	Palustrine Forested Wetland*	6	Wetlands
92	Palustrine Scrub/Shrub Wetland*	6	Wetlands
93	Estuarine Forested Wetland*	6	Wetlands
94	Estuarine Scrub/Shrub Wetland*	6	Wetlands
95	Emergent Herbaceous Wetlands	6	Wetlands
96	Palustrine Emergent Wetland (Persistent)*	6	Wetlands
97	Estuarine Emergent Wetland*	6	Wetlands
98	Palustrine Aquatic Bed*	6	Wetlands
99	Estuarine Aquatic Bed*	6	Wetlands

Canopy

Code	Description
0	0 - 20%
1	21 - 40%
2	41 - 60%
3	61 - 80%
4	81 - 100%

Aspect

Degree Direction	Group
337.5 - 22.5	N
22.5 - 67.5	NE
67.5 - 112.5	E
112.5 - 157.5	SE
157.5 - 202.5	S
202.5 - 247.5	SW
247.5 - 292.5	W
292.5 - 337.5	NW
-1	Flat